

dr hab. TOMASZ PUZYN
profesor nadzwyczajny

Gdańsk, 22 listopada 2016 r.

RECENZJA

rozprawy doktorskiej mgr. Jacka Bogocza

pt. „Eksploracja baz danych w poszukiwaniu nowych reguł projektowania leków”

Przedstawiona mi do recenzji rozprawa doktorska mgr. Jacka Bogocza poświęcona jest nowoczesnym metodom projektowania leków. Ma charakter interdyscyplinarny i w niezwykle ciekawy sposób łączy elementy chemii, farmacji, chemoinformatyki, scientometrii i ekonomii. Praca doktorska wykonana została w Instytucie Chemii na Wydziale Matematyki, Fizyki i Chemii Uniwersytetu Śląskiego pod kierunkiem prof. dr. hab. inż. Jarosława Polańskiego. Warto w tym miejscu podkreślić, że liczne osiągnięcia naukowe zespołu profesora Polańskiego w zakresie chemoinformatyki są wysoko cenione w kraju i za granicą.

Praca liczy 130 stron (bez załączników), zawiera 51 rysunków, 20 tabel oraz odwołuje się do 145 pozycji literaturowych. Wśród załączników umieszczone zostały trzy prace naukowe opublikowane w *Scientometrics* ($IF_{2015}=2,084$), *Drug Discovery Today* ($IF_{2015}=5,625$) oraz *Journal of Computer-Aided Molecular Design* ($IF_{2015}=3,199$).

Tytuł rozprawy odpowiada zawartym w niej treściom, chociaż – ze względu na retrospektywny charakter badań – bardziej poprawnie brzmiałby: „Eksploracja baz danych w celu identyfikacji trendów w zakresie projektowania leków”. Po stronie tytułowej, zwyczajowych podziękowaniach i spisie treści następuje prezentacja problematyki naukowej i celu pracy. Autor zwraca uwagę na złożoność procesu projektowania nowych produktów farmaceutycznych oraz na fakt, że pojęcie leku stanowi obecnie kategorię rynkową. W tym sensie czysto chemoinformatyczne podejście nie wystarcza, a proces wirtualnego projektowania nowych substancji aktywnych nie powinien być prowadzony w oderwaniu od analizy efektów rynkowych. Dlatego głównym celem pracy była próba eksploracji danych chemicznych w powiązaniu z informacjami, które pozwoliłyby wytłumaczyć rolę i wpływ rynku na proces projektowania nowych leków i wprowadzania ich na rynek. To bardzo oryginalne podejście i pierwsza tego rodzaju praca, którą miałem przyjemność recenzować.

W kolejnym rozdziale zatytułowanym „Wstęp” Autor omawia podstawowe koncepcje wykorzystane w pracy, podkreśla wagę interdyscyplinarnego podejścia do problemu oraz streszcza zawartość kolejnych rozdziałów. To bardzo przydatna część pracy, pozwalająca zorientować się czytelnikowi w strukturze i powiązaniach pomiędzy poszczególnymi elementami, które złożyły się na rozprawę.

Kolejne rozdziały (od 3. do 6.) stanowią „Część literaturową” pracy. W rozdziale 3. mgr Bogocz szczegółowo opisuje proces badań nad nowym lekiem. Posługuje się tutaj przykładem procedur wymaganych przez amerykańską Agencję Żywności i Leków (ang. *Food and Drugs Administration, FDA*) argumentując, że procedury wymagane przez FDA stanowią wyznacznik jakości i wzór dla innych podobnych organizacji. Interesujące byłoby również krytyczne porównanie najważniejszych elementów tych procedur z wymogami obowiązującymi w Polsce i w innych krajach Wspólnoty Europejskiej. Proponuję, żeby Kandydat przedstawił takie porównanie w części dyskusyjnej publicznej obrony. Zwracam także uwagę na pewne nieścisłości, które znalazłem w tej części pracy:

- Pierwszy interfejs graficzny zastosowany został w komputerze Xerox (1973 r.), a następnie rozwinięty przez firmę Apple i zastosowany w pierwszym komputerze Macintosh (1984 r.). Wspomniany przez Doktoranta system Windows 1.0 (1985 r.) był jedynie kolejnym graficznym systemem operacyjnym.
- Na stronie 12. znalazłem następujące zdanie: „Po pomyślnych badaniach przedklinicznych *in vitro* związku biologicznie aktywne kieruje się do testów klinicznych *in vivo*.” Czy w fazie przedklinicznej nie przeprowadza się testów *in vitro*? Prosiłbym Doktoranta przy okazji o wyjaśnienie podczas publicznej obrony różnicy pomiędzy badaniami *in vitro* i *in vivo*.
- Omawiając współczynnik podziału $\log P$ (str. 17.) Autor pracy nie odnosi tego współczynnika do termodynamicznego stanu równowagi. Czy Doktorant potrafiłby podać przykład substancji chemicznej, w odniesieniu do której $\log P$ nie byłby odpowiednią miarą lipofilowości, ze względu na niemożność ustalenia się równowagi termodynamicznej i stałego stosunku stężeń pomiędzy stężeniami substancji w modelu fazy lipidowej oktanolu) i wodzie?
- Omawiając wykorzystanie energii najwyższego obsadzonego orbitalu molekularnego (HOMO) i najniższego nieobsadzonego orbitalu molekularnego (LUMO) jako deskryptorów struktury chemicznej związku Doktorant nie wspominał, w jaki sposób się je wyznacza. Proszę o krótką informację na ten temat podczas publicznej obrony.

W rozdziale 4. mgr Bogocz charakteryzuje metody chemii kombinatorycznej *in silico*, ze szczególnym uwzględnieniem algorytmów RECAP, BRICS oraz SIMPLEX. Ta ostatnia nazwa może być mylona z chemoinformatycznym algorytmem optymalizacji simpleksowej, dlatego – w odniesieniu do systemu deskryptorów opracowanych przez grupę prof. Kuz'mina – sugeruję używanie bardziej aktualnego akronimu SiRMS (ang. *Simplex Representation of Molecular Structure*). Ta część napisana jest bardzo dobrze, a idee działania poszczególnych algorytmów przedstawione przejrzysto. Drobną uwagę dotyczy niefortunnego sformułowania na str. 39 dotyczącego metody SIMPLEX. Autor pisze, że jest to „metoda generowania fragmentów molekularnych o liczbie atomowej cztery”. Chodzi tutaj oczywiście o czteroatomowe fragmenty molekularne.

W rozdziale 5. Doktorant omawia specjalistyczne narzędzia wykorzystywane w chemoinformatyce skupiając się na językach programowania, metodach uzyskiwania tzw. informacji chemicznej (kodowania informacji o strukturze chemicznej badanych związków) oraz wykorzystaniu bibliotek chemoinformatycznych. Do tej części nie mam żadnych zastrzeżeń.

Rozdział 6. poświęcony jest metodom scjentometrycznym. Autor koncentruje swoją uwagę na liczbie cytowań, indeksie Hirscha i współczynniku wpływu (ang. *impact factor*, IF) czasopism naukowych. Są to niewątpliwie najważniejsze i najpowszechniej wykorzystywane wskaźniki. Mają jednak one pewne ograniczenia. Na przykład indeks Hirscha nie uwzględnia etapu kariery naukowej ocenianego naukowca. Czy Doktorant potrafiłby podać przykład modyfikacji indeksu Hirscha, której celem jest wyjście poza to ograniczenie?

Niezależnie od wymienionych powyżej nieścisłości, szczegółowa analiza „Części literaturowej” rozprawy potwierdza ogólną wiedzę teoretyczną Kandydata w dyscyplinie i dobre przygotowanie do prowadzenia pracy naukowej, w tym krytyczne podejście do cytowanych pozycji piśmiennictwa.

Kolejne części rozprawy (rozdziały od 7. do 12.) stanowi prezentacja wyników i wniosków płynących z badań własnych. Jako najważniejsze osiągnięcia Doktoranta przedstawione w rozprawie wymieniłbym:

- Wykazanie, że sukces grup naukowych zajmujących się między innymi projektowaniem nowych leków skorelowany jest w znacznym stopniu z poziomem finansowania nauki. To ciekawy i praktyczny wniosek (poparty rzetelną analizą!), który zapewne zainteresowałby zespoły pracujące obecnie nad założeniami reformy szkolnictwa wyższego w Polsce.

- Zaproponowanie koncepcji „starzenia się” leku będącego na rynku oraz metodyki badania tego procesu, a także wskazanie powiązań pomiędzy spadkiem produktywności przemysłu farmaceutycznego a „starzeniem się” leków. Interesującym wnioskiem jest to, że średni wiek leków obecnych na rynku stale rośnie. Wprowadzane na rynek nowe leki nie wykazują dostatecznej efektywności ekonomicznej by wyprzeć więcej niż 3% starszych farmaceutyków z tzw. „listy bestsellerów”.
- Zbadanie wpływu parametrów struktury chemicznej opisywanych tzw. „Regułą 5” wg. Lipińskiego, parametru ADMET wg. Gleesona, logarytmicznej wartości stężenia hamującego leku pIC_{50} na sukces rynkowy leku.
- Przeprowadzenie analizy fragonomicznej, w wyniku której zidentyfikowane zostały fragmenty struktury chemicznej występujące najczęściej w strukturze chemicznej najlepiej sprzedających się leków.

Wymienione osiągnięcia naukowe spełniają zwyczajowe i formalne wymagania stawiane rozprawom doktorskim. Ta część pracy stanowi w mojej ocenie bardzo istotny wkład do współczesnego rozumienia złożoności procesu chemoinformatycznego projektowania leków.

Z funkcji recenzenta wynika również obowiązek wskazania słabszych fragmentów tej części pracy oraz zagadnień, które chciałbym przedyskutować z Kandydatem podczas publicznej obrony.

Po pierwsze, zabrakło mi bardziej wnikliwej dyskusji o wpływie jakości naukowej badań prowadzonych przez poszczególne instytucje badawcze na prawdopodobieństwo sukcesu rynkowego nowego leku. Taka dyskusja stanowić mogłaby pomost pomiędzy częścią bibliometryczną a chemoinformatyczną badań własnych.

Po drugie, ciekawym uzupełnieniem przeprowadzonych badań byłoby pokazanie przynależności poszczególnych leków do grup związków chemicznych wyróżnianych z uwagi na pewne szczególne grupy funkcyjne zdefiniowane *a priori* (np. wiązanie peptydowe). Dodatkowo można byłoby pokazać to w perspektywie czasowej. W ten sposób możliwe byłoby stwierdzenie czy jakaś szczególna grupa związków (np. peptydy i białka) stanowi szczególny obiekt zainteresowania przemysłu farmaceutycznego w ostatnich latach. Takich wniosków nie można niestety wyciągnąć wyłącznie na podstawie znajomości masy cząsteczkowej czy stwierdzenia obecności pewnych dość powszechnie występujących motywów strukturalnych (np. podstawnik fenyłowy, czy kilkuwęglowy łańcuch alifatyczny).

Po trzecie, definicja stężenia hamującego leku (plC_{50}) przedstawiona na stronie 81. nie jest poprawna. Wartość plC_{50} odnosi się do stężenia powodującego zahamowanie aktywności o połowę, nie zaś do „50% wartości maksymalnego stężenia” jak pisze Autor. Z lektury pracy nie wynika również, jakie jednostki stosował Autor dla plC_{50} . Czy były to jednostki stężenia masowego (np. mg/dm^3), czy też stężenia molowego (np. mol/dm^3). Oczywiście to drugie podejście byłoby prawidłowe w kontekście badań przeprowadzonych przez Doktoranta.

Wymienione powyżej niewielkie uchybienia nie umniejszają jednak mojej wysokiej oceny merytorycznej przeprowadzonych badań.

Obowiązkiem recenzenta jest również ocena formalnej strony dysertacji. Rozprawa jest napisana przejrzysto, przy użyciu poprawnego języka. Wyjątkiem jest zbyt dosłowne tłumaczenie niektórych pojęć z języka angielskiego. Przykładem może być zbyt dosłowne tłumaczenie pojęcia „high throughput screening”. Poprawna wersja (zresztą podana przez samego Autora na str. 25.) to „wysokowydajne badania przesiewowe”. Tekst jest sformatowany poprawnie, co czyni pracę bardzo czytelną. Zauważyłem jedynie niewielkie potknięcia w tym zakresie:

- Opis danych na stronach 61-63 nie zgadza się z zawartością tabeli w załączniku 1. W tekście pracy na str. 61 znajduje się również błędnie podany odnośnik (powinno być „w tabeli 18.1”).
- Co oznaczają kolory (niebieski i pomarańczowy) na rysunku 8.2.?
- Brak jednostek w tabeli 10.5.

Z załączonego do rozprawy wykazu dorobku naukowego dowiedziałem się, że mgr Jacek Bogocz – oprócz autorstwa trzech publikacji w czasopismach z listy JCR – jest autorem czterech patentów, dwóch rozdziałów w książkach oraz zaprezentował swoje wyniki aż na trzynastu krajowych i międzynarodowych konferencjach naukowych. Dorobek naukowy Kandydata łączy więc elementy badań podstawowych i przemysłowych, a biegłość w prezentowaniu wyników prac badawczych jest – oprócz prowadzenia samych badań – ważną umiejętnością pożądaną w świecie nauki i świadczy o bardzo dobrym przygotowaniu Doktoranta do pracy naukowej.

W mojej ocenie rozprawa stanowi oryginalne rozwiązanie kilku bardzo istotnych problemów naukowych oraz potwierdza ogólną wiedzę teoretyczną Kandydatki w dyscyplinie, a także umiejętność samodzielnego prowadzenia pracy naukowej. Drobne potknięcia, które wymieniłem w niniejszej recenzji nie wpływają na wysoką ocenę merytoryczną przeprowadzonych badań. Dlatego z pełnym przekonaniem

stwierdzam, że rozprawa mgr. Jacka Bogocza spełnia wymagania określone w Art. 13. *Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki* z dnia 14 marca 2013 r. z późniejszymi zmianami. Wniosuję zatem o dopuszczenie Kandydata do dalszych etapów przewodu doktorskiego.



KIEROWNIK
Pracowni Chemometrii Środowiska
dr hab. Tomasz Puzyn, prof. UG

Tomasz Puzyn